

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-334049

(P2002-334049A)

(43)公開日 平成14年11月22日 (2002. 11. 22)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テマコード <sup>*</sup> (参考)
G 0 6 F 13/12	3 3 0	G 0 6 F 13/12	3 3 0 F 5 B 0 0 5
3/06	3 0 1	3/06	3 0 1 R 5 B 0 1 4
	3 0 4		3 0 4 F 5 B 0 1 8
12/00	5 1 4	12/00	5 1 4 A 5 B 0 6 5
	5 3 1		5 3 1 D 5 B 0 8 2

審査請求 未請求 請求項の数5 O L (全 12 頁) 最終頁に続く

(21)出願番号 特願2001-140954(P2001-140954)

(22)出願日 平成13年5月11日(2001. 5. 11)

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 藤本 健雄

神奈川県小田原市中里322番地2号 株式会社日立製作所RAIDシステム事業部内

(72)発明者 木城 茂

神奈川県小田原市中里322番地2号 株式会社日立製作所RAIDシステム事業部内

(74)代理人 100080001

弁理士 筒井 大和

最終頁に続く

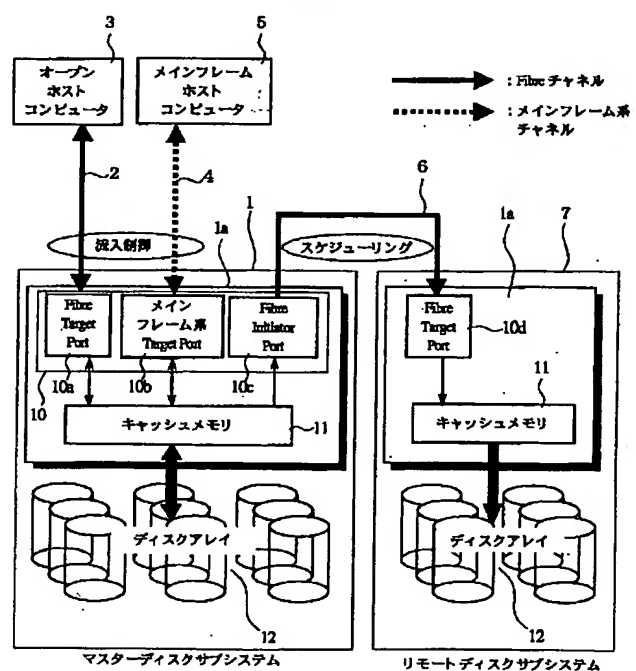
## (54)【発明の名称】 記憶サブシステムおよび記憶サブシステムの制御方法

## (57)【要約】

【課題】 特定の上位装置によるデータ書込負荷の増大に影響されずに、複数の上位装置の配下の記憶サブシステム間で効率的なリモートコピーを実現する。

【解決手段】 入出力ポート10を介してオープンホストコンピュータ3およびメインフレームホストコンピュータ5、リモートディスクサブシステム7に接続されるマスターディスクサブシステム1において、キャッシュメモリ11内に設けられ、リモートコピーに用いられるサイドファイルにおける各ホスト毎の占有データ量の比率を個別に制御する個別比率閾値、および合計比率閾値を設定し、サイドファイルの使用量の合計が合計比率閾値を超過した時点で、各ホストによるサイドファイルへの書込データ量が個別比率閾値を超えないように、各ホストからのライトデータの流入量を制限し、個別比率閾値に応じた割合でリモートディスクサブシステム7へのリモートコピーを実行する。

図 1



## 【特許請求の範囲】

【請求項 1】 複数の上位装置に第 1 のインタフェースを介して接続される第 1 の記憶サブシステムと、第 2 のインタフェースを介して前記第 1 の記憶サブシステムに接続される第 2 の記憶サブシステムとを含み、前記上位装置から前記第 1 の記憶サブシステムに書き込まれるライトデータを、前記第 1 の記憶サブシステムから前記第 2 の記憶サブシステムに複写することで、前記第 1 および第 2 の記憶サブシステムにて前記ライトデータを多重に保持する記憶サブシステムであって、前記第 1 の記憶サブシステムは、個々の前記上位装置毎に、当該第 1 の記憶サブシステム内のデータバッファにおける前記第 2 の記憶サブシステムへの複写が未完の前記ライトデータの占有比率を制御する第 1 の閾値と、個々の前記上位装置毎に設定された前記第 1 の閾値に基づいて個々の前記上位装置からのデータ書込要求の処理を遅延させる制御論理と、を備えたことを特徴とする記憶サブシステム。

【請求項 2】 請求項 1 記載の記憶サブシステムにおいて、

前記第 1 の記憶サブシステムは、複数の前記上位装置による前記データバッファにおける、前記第 2 の記憶サブシステムへの複写が未完の前記ライトデータの総和の占有比率を制御する第 2 の閾値を備え、前記制御論理は前記第 2 の閾値に基づいて、前記第 1 の閾値による前記上位装置からのデータ書込要求の処理を遅延させる動作の開始契機を制御する機能を備えたことを特徴とする記憶サブシステム。

【請求項 3】 請求項 1 記載の記憶サブシステムにおいて、

前記第 1 のインタフェースは、メインフレーム系ホストインタフェースおよびオープン系ホストインタフェースの少なくとも一方からなり、前記第 2 のインタフェースは、ファイバチャネルからなることを特徴とする記憶サブシステム。

【請求項 4】 複数の上位装置に第 1 のインタフェースを介して接続される第 1 の記憶サブシステムと、第 2 のインタフェースを介して前記第 1 の記憶サブシステムに接続される第 2 の記憶サブシステムとを含み、前記上位装置から前記第 1 の記憶サブシステムに書き込まれるライトデータを、前記第 1 の記憶サブシステムから前記第 2 の記憶サブシステムに複写することで、前記第 1 および第 2 の記憶サブシステムにて前記ライトデータを多重に保持する記憶サブシステムの制御方法であって、個々の前記上位装置毎に、前記第 1 の記憶サブシステム内のデータバッファにおける前記第 2 の記憶サブシステムへの複写が未完の前記ライトデータの占有比率を制御する第 1 の閾値を設定する第 1 のステップと、個々の前記上位装置毎に設定された前記第 1 の閾値に基づいて個々の前記上位装置からのデータ書込要求の処理

を遅延させる第 2 のステップと、

を実行することことを特徴とする記憶サブシステムの制御方法。

【請求項 5】 請求項 4 記載の記憶サブシステムの制御方法において、

前記第 1 のステップでは、複数の前記上位装置による前記データバッファにおける、前記第 2 の記憶サブシステムへの複写が未完の前記ライトデータの総和の占有比率を制御する第 2 の閾値を設定する操作が行われ、前記第 2 のステップでは、前記第 2 の閾値に基づいて、前記第 1 の閾値による前記上位装置からのデータ書込要求を待たせる操作の開始契機の制御を行うことを特徴とする記憶サブシステムの制御方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、記憶サブシステムおよびその制御技術に関し、特に、遠隔地等に設置された複数の記憶サブシステム間でのデータ複写によるデータの多重化技術等に適用して有効な技術に関する。

## 【0002】

【従来の技術】昨今の情報化社会の進展により、社会活動の各分野での情報処理に用いられるコンピュータシステムの記憶装置に格納するデータ量が爆発的に増え、また、データとシステムの両方に対して高い信頼性が要求されている。データの信頼性とシステム動作の保証を同時に満たす手法として、記憶装置自体の多重化が行われている。ホストコンピュータから出力されるデータは、直接接続されている記憶装置にだけでなく、その記憶装置を経由して他の記憶装置へもコピーされる。

【0003】一般的に、この手法をリモートコピーと称し、高信頼性が要求されるコンピュータシステムへの適用が行われている。このリモートコピー技術によれば、一方の記憶装置に障害が発生して、動作不可能な状態に陥っても、他方の記憶装置のデータを用いてシステム動作が可能である。

【0004】ホストコンピュータが、当該ホストコンピュータと接続されている記憶装置（以下、マスター側記憶装置と記す）へデータを格納する際、そのデータ転送と同期、又は非同期に、マスター側記憶装置は、マスター側記憶装置に接続される他の記憶装置（以下、リモート側記憶装置と記す）へ、当該データをコピーできるリモートコピーの制御方式が実現されている。

【0005】マスター側記憶装置に障害が発生して動作不可となった場合は、アクティブとなる記憶装置をリモート側に切り替えるだけで、リモート側記憶装置に残っているデータがそのまま利用可能となる。

【0006】また、公衆回線等を経由して、遠隔地へのリモートコピーも行われている。マスター側記憶装置が設置されている場所や都市に大規模な天災等が発生しても、遠隔地にあるリモート側記憶装置に被害が及ばない

限り、リモート側システムへの切り替えを行うだけで、システム全体としては継続動作が可能となる。

【0007】一般的に、遠隔地へのリモートコピーはデータ転送に時間がかかるため、マスター側記憶装置はデータが内部バッファに格納された時点で、ホストコンピュータへ受領完了の応答を行い、その後、ホスト転送とは非同期にリモート側記憶装置へのデータ転送を行う。

【0008】現在では、上記技術はデータの入出力が行われている最中にも実現され、オンラインシステムを停止せずにリモート側へのコピー開始、またはシステム切り替えが可能となっている。

【0009】

【発明が解決しようとする課題】近年、ファイバチャネル・プロトコル(Fibre Channel Protocol)がANSI, NCITS T11で標準化されている等、ホストコンピュータとマスター側記憶装置間の転送性能が飛躍的に向上し続けている。その一方、マスター側記憶装置とリモート側記憶装置間の接続チャネル資源は経済的な理由に制限され、無限に広げることができない。特に、リモートコピー機能を有するシステムを構築する際、広域災害への対応を考えれば、リモート側記憶装置をマスター側記憶装置から遠く離して設置したいが、遠隔地へのリモートコピーを行うことになると、公衆回線を利用しても、代りに専用線を敷設しても、膨大な費用が必要となってしまう。

【0010】その結果、ホストコンピュータとマスター側記憶装置間のデータ転送チャネルと比較して、マスター側記憶装置とリモート側記憶装置間のリモートコピーチャネルは、どうしても本数が制限され、データ転送の限界性能も比較的に低いものとなる。

【0011】従って、例えば、ある時間帯にマスター側各ホストコンピュータからのデータ格納が、リモートコピーチャネルの性能限界を超えて大量に発生すると、リモートコピーが追従できずに、マスター側記憶装置内のデータバッファに未コピーのデータが溢れてしまう。マスター側記憶装置のデータバッファが満杯になると、データロストが発生しない様に、ホストから新たなデータ書き込みを受け付けなくするが、マスター側記憶装置に接続されている複数ホストコンピュータの内、1つのホストから大量なデータが流れてくるだけで、他のホストコンピュータからもデータ格納ができなくなってしまう。

【0012】マスター側記憶装置に複数ホストコンピュータが接続されている場合、全てのホストチャネルが同じ重要度を持っているとは限らない。各ホストチャネルの性能要求や転送されるデータ量にもばらつきがあると考えられる。例えば、性能要求が低く、もしくは比較的に重要でないホストチャネルからのデータ格納が多量に発生し、上述のデータバッファ満杯の現象を引き起こしても、性能要求が高かったり、重要だったりする他のホ

ストチャネルからもデータの書き込みができなくなる不具合が考えられ、運用上の大きな技術的課題となっている。

【0013】一方、近年では、小型コンピュータ用の汎用OSを搭載したパーソナルコンピュータあるいはワークステーション等の性能向上に伴い、ホストコンピュータとして、メインフレーム以外に、これらの小型コンピュータもホストコンピュータとして記憶サブシステムに接続されるに至っている。

【0014】このため、記憶サブシステム側では、メインフレーム系ホストインタフェースの他に、小型コンピュータ用の、いわゆるオープン系ホストインタフェースも備えて、メインフレーム系およびオープン系のホストコンピュータの双方に記憶サブシステムを共有させる運用形態も考えられる。

【0015】その場合、上述のようにリモートコピーチャネルにおける経済的な制約により、データのリモートコピーを行う複数の記憶サブシステムを、上述の多様なインタフェースにて接続することは現実的ではなく、異なるホスト接続インタフェース間におけるデータフォーマットの統一が必要となる、という技術的課題もある。

【0016】本発明の目的は、記憶サブシステムに接続される複数の上位装置の各々の重要度に応じた優先度にて、複数の記憶サブシステム間でのライトデータのリモートコピーを行うことが可能な記憶サブシステムおよびその制御技術を提供することにある。

【0017】本発明の他の目的は、記憶サブシステムに接続される複数の上位装置のうち、特定の上位装置によるデータ書込負荷の増大に影響されることなく、複数の記憶サブシステム間での効率的なリモートコピーを実現することが可能な記憶サブシステムおよびその制御技術を提供することにある。

【0018】本発明の他の目的は、複数の上位装置の各々と記憶サブシステム間の多様な接続インタフェースを、データのリモートコピーを行う複数の記憶サブシステム間の接続インタフェースに統一することで、効率的なリモートコピーシステムの構築が可能な技術を提供することにある。

【0019】

【課題を解決するための手段】本発明は、複数の上位装置に第1のインタフェースを介して接続される第1の記憶サブシステムと、第2のインタフェースを介して第1の記憶サブシステムに接続される第2の記憶サブシステムとを含み、上位装置から第1の記憶サブシステムに書き込まれるライトデータを、第1の記憶サブシステムから第2の記憶サブシステムに複写することで、第1および第2の記憶サブシステムにてライトデータを多重に保持する記憶サブシステムであって、第1の記憶サブシステムは、個々の上位装置毎に、当該第1の記憶サブシステム内のデータバッファにおける第2の記憶サブシ

ムへの複写が未完のライトデータの占有比率を制御する第1の閾値と、個々の上位装置毎に設定された第1の閾値に基づいて個々の上位装置からのデータ書込要求を待たせる制御論理と、を備えたものである。

【0020】また、本発明は、複数の上位装置に第1のインタフェースを介して接続される第1の記憶サブシステムと、第2のインタフェースを介して第1の記憶サブシステムに接続される第2の記憶サブシステムとを含み、上位装置から第1の記憶サブシステムに書き込まれるライトデータを、第1の記憶サブシステムから第2の記憶サブシステムに複写することで、第1および第2の記憶サブシステムにてライトデータを多重に保持する記憶サブシステムの制御方法であって、個々の上位装置毎に、第1の記憶サブシステム内のデータバッファにおける第2の記憶サブシステムへの複写が未完のライトデータの占有比率を制御する第1の閾値を設定する第1のステップと、個々の上位装置毎に設定された第1の閾値に基づいて個々の上位装置からのデータ書込要求を待たせる第2のステップと、を実行するものである。

【0021】より具体的には、本発明では、記憶サブシステム間のリモートコピーにおいて、リモートコピーの経路としてファイバチャネル・プロトコルを採用するとともに、以下の手段を用いる。

【0022】(1) 第1の記憶サブシステムであるマスター側記憶装置に接続される各ホスト(上位装置)単位に、使用できるバッファ容量等の資源の比率を定義し、ユーザーインタフェース経由で設定できる手段を設ける。

【0023】(2) マスター側記憶装置内のデータバッファの使用容量を、各ホスト単位に個別に自動に計測し、各ホストが使用しているバッファ容量の比率を求める。

【0024】(3) データバッファの合計使用量が高い場合、各ホストが使用しているバッファ容量の比率を、前記ユーザー設定で定めた使用できる資源の比率と比較し、使用量が多過ぎたホストチャネルからのライトデータの流入を制限する。

【0025】(4) ホストチャネルからのデータ流入制限手段として、マスター側記憶装置にホストを待たせる論理を持たせる。

【0026】(5) マスター側記憶装置内のデータバッファから、第2の記憶サブシステムであるリモート側記憶装置へのデータコピーはランダムに行う。その結果、各ホストコンピュータ側から見ると、設定された比率の通りにリモートコピーのスケジューリングが行われることになる。

【0027】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【0028】図1は、本発明の一実施の形態である記憶

サブシステムの制御方法を実施する記憶サブシステムを用いた情報処理システムの構成の一例を示す概念図であり、図2は、本実施の形態の記憶サブシステムの内部構成の一例をより詳細に例示した概念図である。

【0029】本実施の形態の記憶サブシステムは、マスターディスクサブシステム1と、リモートディスクサブシステム7と、これらを接続するリモートコピー用ファイバチャネル6で構成されている。

【0030】マスターディスクサブシステム1は、外部に接続する複数の入出力ポート10と、入出力データを一時的に格納するバッファの役割をも果たすキャッシュメモリ11、共用メモリ11-1等を含む記憶制御装置1aと、データ記憶媒体であるディスクアレイ12から構成される。マスターディスクサブシステム1の入出力ポート10を構成するファイバチャネルターゲットポート10aは、ファイバチャネル2を経由して、オープンホストコンピュータ3と接続し、メインフレーム系チャネルターゲットポート10bは、メインフレーム系チャネル4を経由して、メインフレームホストコンピュータ5と接続されている。

【0031】また、マスターディスクサブシステム1の入出力ポート10を構成するファイバチャネルイニシエータポート10cは、リモートコピー用ファイバチャネル6を経由してリモートディスクサブシステム7のファイバチャネルターゲットポート10dに接続する。すなわち、マスターディスクサブシステム1は、リモートディスクサブシステム7に対してはホストコンピュータとして振る舞う。

【0032】マスターディスクサブシステム1の記憶制御装置1aには、キャッシュメモリ11とディスクアレイ12との間におけるデータの入出力を制御する複数のディスクアダプタ12aが設けられている。

【0033】記憶制御装置1aにおけるファイバチャネルターゲットポート10a、メインフレーム系チャネルターゲットポート10b、ファイバチャネルイニシエータポート10c、およびディスクアダプタ12a、さらには、キャッシュメモリ11および共用メモリ11-1は、バス13を介して相互に接続され、このバス13を介して、これらの各要素間での情報の授受が行われる。

【0034】本実施例ではバス13を用いて説明しているが、各要素の接続手段としては、スイッチを用いたスターネット接続で実現してもよい。

【0035】すなわち、ファイバチャネルターゲットポート10a、メインフレーム系チャネルターゲットポート10b、ファイバチャネルイニシエータポート10c、およびディスクアダプタ12aは、各々が独立したプロセッサ(CPU)を備え、共用メモリ11-1に設定される後述のような制御情報の参照および更新にて、並行動作により、データの入出力処理等における相互に協調した制御動作を行う。

【0036】リモートディスクサブシステム7の構成は、マスターディスクサブシステム1と等価であるため、マスターディスクサブシステム1と同一機能部分は、共通の符号を付して重複した説明は割愛する。

【0037】なお、図1に例示された本実施の形態のマスターディスクサブシステム1のファイバチャネルターゲットポート10a、メインフレーム系チャネルターゲットポート10b、ファイバチャネルイニシエータポート10cは、1つずつしか示していないが、それぞれ複数個に増設できる。また、リモートディスクサブシステム7の入出力ポートも同様である。

【0038】リモートディスクサブシステム7には、マスターディスクサブシステム1と同様に図示しないホストコンピュータが接続されていてもよい。

【0039】図2に例示されるように、マスターディスクサブシステム1は、LAN等の外部接続インタフェース15を持ち、小型パーソナルコンピュータ等で構成されるサービスプロセッサ14（SVP）が接続されている。ユーザーは、SVP14を通じてマスターディスクサブシステム1へリモートコピーのパス構成を定義できる他、後述のような本実施の形態における後述のサイドファイルの使用に関わる各種制御情報として、システムに合計比率閾値40、各ターゲットポート毎に個別比率閾値32を設定することも可能となっている。

【0040】図3は、本実施の形態における記憶サブシステムにて用いられる制御情報の一例を示す概念図である。

【0041】本実施の形態の場合、共用メモリ11-1には、制御情報としてサイドファイル管理テーブル30、合計比率閾値40、サイドファイル容量41、キャッシュ管理テーブル50が設定される。

【0042】サイドファイル管理テーブル30は、オープンホストコンピュータ3、メインフレームホストコンピュータ5と接続されるファイバチャネルターゲットポート10a、メインフレーム系チャネルターゲットポート10b、等のホストコンピュータ接続ポート（すなわちホストコンピュータそのもの）を個別に識別するための複数のポート番号31の各々のエントリ毎に、当該ポートに接続されるホストコンピュータ毎に設定される個別比率閾値32、サイドファイル内の個別使用量33等の情報が設定される。

【0043】キャッシュ管理テーブル50には、キャッシュメモリ11の記憶領域を分割して設定された複数のセグメントを識別するためのセグメント番号51と、このセグメント番号51の各々に対応して設けられたマスターダティフラグ52、リモートダティフラグ53、LRU管理情報54、等が格納されている。

【0044】キャッシュメモリ11は上述のセグメントを単位としてデータの格納が行われ、ファイバチャネルターゲットポート10aおよびメインフレーム系チャネ

ルターゲットポート10bによるホスト側からキャッシュメモリ11へのデータ書込に際しては、キャッシュメモリ11のセグメントに格納された時点で、書込完了がホストに应答され、ディスクアダプタ12aによるキャッシュメモリ11からディスクアレイ12への実際の書込、およびファイバチャネルイニシエータポート10cによるキャッシュメモリ11からリモートディスクサブシステム7へのリモートコピーは、ホスト側からキャッシュメモリ11へのデータ書込とは非同期に実行される。

【0045】すなわち、マスターダティフラグ52（＝1）は、ホスト側からキャッシュメモリ11に書き込まれたライトデータを保持するセグメントのうち、当該ライトデータがディスクアレイ12側に未反映の状態を示し、ファイバチャネルターゲットポート10aおよびメインフレーム系チャネルターゲットポート10bによってデータ書込時にセット（＝1）され、ディスクアダプタ12aにて、ディスクアレイ12への反映後にリセット（＝0）される。

【0046】リモートダティフラグ53（＝1）は、ホスト側からキャッシュメモリ11に書き込まれたライトデータを保持するセグメントのうち、リモートディスクサブシステム7側へのリモートコピーが未完の状態を示し、ファイバチャネルターゲットポート10aおよびメインフレーム系チャネルターゲットポート10bによってキャッシュメモリ11へのライトデータの書込時にセット（＝1）され、リモートコピー完了後に、ファイバチャネルイニシエータポート10cによってリセット（＝0）される。

【0047】本実施の形態の場合、サイドファイルとは、キャッシュメモリ11上の上述のセグメントにおいて、リモートダティフラグ53（＝1）がセットされ、リモートコピー未完のデータを一時保持した状態にあるバッファ（セグメント）の集合を示す。

【0048】上述のサイドファイル容量41は、このサイドファイルを構成するセグメントの最大数を示し、システム管理者によって、SVP14から設定される。

【0049】LRU管理情報54は、LRU等のアルゴリズムによって、新たなライトデータに対するキャッシュメモリ11のセグメントの割り当てや、用済みのセグメントの解放を制御する情報が格納される。

【0050】本実施の形態において、合計比率閾値40とは、サイドファイルの使用済み容量に対して、流入制限制御を開始する契機を与える閾値である。リモートディスクサブシステム7へのリモートコピーが完了していないデータの総和（個別使用量33の合計）である合計使用量34が占めるサイドファイルのサイドファイル容量41に対する比率が本閾値を越えた場合に、後述の個別比率閾値32を参照してのホスト側からのライトデータの流入制限を開始する。

【0051】サイドファイル管理テーブル30における個別比率閾値32とは、個々のホストから格納されるデータが、それぞれ使用できるサイドファイル容量の比率であり、後で詳細を述べるが、リモートディスクサブシステム7へのデータコピーも本比率で定義されている優先度で行われることになる。デフォルト値は、接続されている各ホストコンピュータが全サイドファイルを均等に使える様に設定されるが、ユーザーは各ホストコンピュータのデータアクセスパターンを加味しながら、アクセス量が多く、性能要求が高く、又は重要なホストチャネルに対して、高い個別比率閾値32を定義することができる。

【0052】例として、ホストコンピュータが4つ接続されるマスターディスクサブシステムにおいて、合計比率閾値と個別比率閾値を設定した一例を模式図として図4に示す。この図4の例では、ホストA～ホストDによるサイドファイル内のデータの総和が、サイドファイル容量41の60%を超過した時点で、個別比率閾値32による、各ホストコンピュータ毎のライトデータの流入制限制御が開始される。

【0053】図4の場合、この流入制限制御の開始以後におけるライトデータの受け付けおよびリモートコピーの優先度は、個別比率閾値32の大きい順に、ホストD、ホストA、ホストB、ホストCとなる。

【0054】以下、本実施の形態の作用の一例について説明する。

【0055】まず、SVP14を介して、合計比率閾値40、サイドファイル容量41、個別比率閾値32を適宜設定する。この設定操作は、稼働中に随時行うこともできる。

【0056】システム立ち上げによる初期化後、又はSVP14からリモートコピーパス形成指示を受けた契機に、ファイバチャネルイニシエータポート10cは、アドレスを検出した全ポートにN\_Portログイン(PLOGI)処理を発行する。ファイバチャネル接続機器は通常PLOGIに対して応答(ACC)を返す際、そのペイロードにファイバチャネル規格に定められているポート固有のWWN(World Wide Name)を格納する。本実施の形態のリモートコピーに対応するポートは予め一定ルールに従ってWWNを定めるので、ファイバチャネルイニシエータポート10cは各ACCに格納されているWWNをチェックすることにより、リモートコピー対象のリモートディスクサブシステム7のファイバチャネルターゲットポート10dを識別できる。

【0057】目的外のポートだと判定したら、直ちにログアウト(LOGO)を出力し、当該ポートからログアウトする。リモートコピーの対象のリモートディスクサブシステム7のファイバチャネルターゲットポート10dだと判定したら、続けてプロセスログイン(PRL

1)を発行し、リモートコピーのパス形成を行う。

【0058】このような準備処理の後、図8に例示されるフローチャートにて、ホスト側からのライトデータの流入制限を伴うライトコマンドの処理を実行する。

【0059】たとえば、オープンホストコンピュータ3から、データライトコマンド(ライトコマンド以外は、ステップ113で処理しコマンド待ちに戻る)を受けたマスターディスクサブシステム1のファイバチャネルターゲットポート10aは(ステップ101、ステップ102)、ライトデータを一時的にキャッシュメモリ11へ格納するとともに(ステップ103)、リモートコピーのパス構成情報をチェックする(ステップ104)。リモートコピーのパスが形成されていなければ、マスターティフラグ52をセットして、そのままデータ受領(ライト完了)の応答をオープンホストコンピュータ3に返すが(ステップ112)、リモートコピーのパスが形成されていれば、続いて合計比率閾値40のチェックを行う(ステップ105)。

【0060】このステップ105で、キャッシュメモリ11上のサイドファイルの合計使用量34(複数のポートの個別使用量33の総和)を計算し、使用済み容量(合計使用量34)のサイドファイル容量41に対する比率が、システムに対して予め設定された合計比率閾値40を超えていなければ(合計使用量34 ≤ サイドファイル容量41 × 合計比率閾値40)、リモートマスターティフラグ53をセットし、当該ライトデータが保持されたセグメントをサイドファイルに追加する(ステップ110)とともに、合計使用量34(個別使用量33の総計)および個別使用量33の更新を行い(ステップ111)、さらにマスターティフラグ52をセットし、データ受領の応答をホストコンピュータに返す(ステップ112)。

【0061】ステップ105で、使用済み容量(個別使用量33の総計)が合計比率閾値40を上回った場合(合計使用量34 > サイドファイル容量41 × 合計比率閾値40)、更に当該ファイバチャネルターゲットポート10aに設定されている個別比率閾値32をチェックする(ステップ106)。

【0062】即ち、このステップ106で、オープンホストコンピュータ3から格納されたライトデータが使用したサイドファイル内の個別使用量33の合計使用量34に対する比率が、当該ファイバチャネルターゲットポート10aに対して、予め設定された個別比率閾値32を超えていなければ(個別使用量33 ≤ 個別比率閾値32 × 合計使用量34)、そのまま正常終了の応答をホストコンピュータに返す(前述のステップ110～ステップ112を実行)が、個別比率閾値32を上回った場合(個別使用量33 > 個別比率閾値32 × 合計使用量34)は、ホストコンピュータの種別を判別し(ステップ107)、この場合オープン系なので、

ファイバチャネルプロトコルに含まれるスリープ処理を実行し（ステップ108）、オープンホストコンピュータ3への応答をできる限り遅らせる。ここで、応答を遅延させるのは、ホスト側からのデータの流入速度を制限するためである。データ流入速度が緩和されている間に、非同期に並行して実行される後述の図9のフローチャートに例示されるリモートディスクサブシステム7へのリモートコピーが進めば、コピーが完了したデータ分のサイドファイルが解放され、データバッファ上に占める当該ホストコンピュータのライトデータの割合が減少していくことは容易に想像できる。

【0063】マスターディスクサブシステム1のファイバチャネルターゲットポート10aが、オープンホストコンピュータ3にデータ受領を応答した後、当該ファイバチャネルターゲットポート10a（オープンホストコンピュータ3）が使用したバッファ容量（個別使用量33）の加算を行い、次の判定に使うために保存する。

【0064】図9のフローチャートに例示されるように、ホスト側からキャッシュメモリ11へのデータ書込とは非同期にリモートコピーを実行するマスターディスクサブシステム1のファイバチャネルイニシエータポート10cには、リモートコピーをスケジューリングするマイクロプログラムが内蔵されている。

【0065】すなわち、サイドファイルの管理情報（リモートダーティフラグ53）を常時検索して（ステップ201、ステップ202）、リモートコピーが完了していないデータ（セグメント）をランダムに選択し（ステップ203）、リモートコピーパスが形成されているリモートディスクサブシステム7のファイバチャネルターゲットポート10dに対して、リモートコピー専用のコマンドを発行する（ステップ204）。

【0066】本実施の形態では、マスターディスクサブシステム1のファイバチャネルイニシエータポート10cと、リモートディスクサブシステム7のファイバチャネルターゲットポート10dには、SCSI-FCPプロトコルに独自のコマンド（Vendor Unique Command）を追加している。SCSI-FCPプロトコルはFCP CDBを16バイトのフォーマットで規定している。リモートコピー専用コマンドは、ライトコマンド、リードコマンド、コントロール/センスコマンドの3種類に大別され、ここで詳細フォーマットは省略するが、いずれも先頭バイトのオペコードに特殊コマンド（Vendor Unique Command）を設定し、以下サブコード、パラメータの各バイトに転送長、サブブロック種別等の詳細情報を格納する。このように、リンクレベルを含めて、ファイバチャネルプロトコルのフレーム形式を遵守しながら、リモートコピー独自の処理を実行可能とする。

【0067】リモートディスクサブシステム7のファイバチャネルターゲットポート10dは、ホストからの通

常コマンド以外、上記リモートコピー専用コマンドにも対応できる制御論理を備えているものとする。リモートディスクサブシステム7のファイバチャネルターゲットポート10dはリモートコピー用のライトコマンドを受領したら、データを格納する領域をキャッシュメモリ上に確保し、転送レディを返す。

【0068】マスターディスクサブシステム1のファイバチャネルイニシエータポート10cは転送レディを受領すると、リモートコピー用ファイバチャネル6を経由してデータ転送を行う。すべてのデータがライト完了した後、最後にファイバチャネルターゲットポート10dからのRSPを受領してコピー動作を完結するとともに、マスターディスクサブシステム1のキャッシュメモリ11におけるリモートコピー済みの当該セグメントのリモートダーティフラグ53をリセット（=0）し、サイドファイルからの削除を実行し（ステップ205）、対応ホストの個別使用量33の減算更新を実行する（ステップ206）。

【0069】上述のマスターディスクサブシステム1と、リモートディスクサブシステム7間のリモートコピーのためのコマンド及びデータのやり取りシーケンスの一例を図5に示す。

【0070】なお、前に述べた様に、マスターディスクサブシステム1のファイバチャネルイニシエータポート10cがコピー転送を行う順序は、サイドファイル内にコピーが完了していないデータ（セグメント）（リモートダーティフラグ53=1のもの）をランダムに選択（ステップ203）して決定されるので、各ホストコンピュータのデータは結果的に、個別比率閾値32の大小に依存した優先順位でリモートコピーされることになる。

【0071】上述の図8のフローチャートにおいて、マスターディスクサブシステム1が、メインフレーム系チャネル4経由で、メインフレームホストコンピュータ5からデータライトコマンドを受けた場合は、メインフレーム系チャネルターゲットポート10bは、リモート側へコピーする前、キャッシュメモリ11に格納するタイミング（ステップ103）でデータフォーマットの変換を行う。

【0072】図6に、この時のデータフォーマットの変換方法の一例を示す。メインフレームホストコンピュータシステム（メインフレームホストコンピュータ5）で使われるCKDデータフォーマットのレコードは、C部（カウント部）、K部（キー部）、D部（データ部）から形成され、レコード内のC部、K部、D部間、及びレコード間には、データを含まないGAPと呼ばれる部分で区切られている。C部、K部、D部はそれぞれ不定長である。オープン系のホストコンピュータで一般に使用されるSCSIの規格ではデータ格納単位であるブロックが固定サイズであり、CKDデータフォーマットのデ

ータをそのままSCSI-FCP準拠のコマンドで転送（リモートコピー）できない。そこで、マスターディスクサブシステム1のメインフレーム系チャネルターゲットポート10bは、ホスト（メインフレームホストコンピュータ5）からのライトデータを受けながら、C部の位置を対応する一定長ブロックの先頭に来るように定め、GAP部分を捨て、後続するK部、D部を前に詰めた上で、キャッシュメモリ11に格納する。また、レコード間には区切りのGAPを埋める。各レコードのC部の位置を対応するブロックの先頭に置くのは、メインフレームホストコンピュータ5からのアクセスが行われた時、目的レコードのサーチが単純計算によって、高速に行える様にするためである。以降、キャッシュメモリ11に格納されているデータは、固定長ブロックに分割されたものとして扱う。キャッシュメモリ11からディスクアレイ12へのデータ格納時にも、リモートコピー用ファイバチャネル6を経由した、リモートディスクサブシステム7へのデータコピー時にも、オープン系データと同じ様に扱う。

【0073】データフォーマット変換処理以外は、マスターディスクサブシステム1からリモートディスクサブシステム7へのリモートコピー動作は共通である。合計比率閾値40及び個別比率閾値32を用いて流入制限を行う制御は、メインフレームホストコンピュータ5の接続時にも適応される。但し、メインフレームホストコンピュータ5とのインタフェース規格上、スリープによる単純な遅延が長く持続できないので、メインフレームホストコンピュータ5を待たせる手段としてコマンドリトライを要求する応答を返し、ステップ103でライトコマンド受領時に一時的にキャッシュメモリ11に保持したデータは破棄することになる（図8のステップ107、ステップ109）。

【0074】マスターディスクサブシステム1のファイバチャネルイニシエータポート10cと、リモートディスクサブシステム7のファイバチャネルターゲットポート10dも、オープンコンピュータ用とメインフレームホストコンピュータ用とを区別せず、同一構造のポートが使用可能である。

【0075】図10のフローチャートにて、キャッシュメモリ11内のライトデータをディスクアレイ12に反映させる処理の一例について説明する。

【0076】ディスクアダプタ12aは、共用メモリ11-1のキャッシュ管理テーブル50を検索し（ステップ301）、ディスクに未反映のデータ（マスタデータフラグ52=1のセグメント）があるか調べ（ステップ302）、ある場合には、たとえば、LRU等の方法で最も古い、すなわち長くキャッシュメモリ11内に留まっていたセグメントを選択して（ステップ303）、当該セグメントのディスクアレイ12への書込を実行（ステップ304）した後、当該セグメントのマスタダ

ータフラグ52をリセット（=0）するとともにLRU管理情報54の更新を行う（ステップ305）。

【0077】本実施の形態では、リモートコピーの経路としてファイバチャネル・プロトコルを用いるので、図7に示す様に、マスタ側からリモート側をファイバチャネルケーブルで直結するルート（R1）と、ファイバチャネルハブ60を経由してリモート側に接続するルート（R2）と、ファイバチャネルスイッチ71からファブリック70（Fabric）に連結し、リモート側に接続するルート（R3）と、ファイバチャネルハブ60を経由してファイバチャネルスイッチ71からFabric70に連結し、リモート側に接続するルート（R4）とが考えられる。これらのルート（R1）～ルート（R4）の接続方式はすべてファイバチャネルプロトコルで規格化されている接続モードであり、前述した様に、本実施の形態の場合には、マスターディスクサブシステム1のファイバチャネルイニシエータポート10cからリモートディスクサブシステム7のファイバチャネルターゲットポート10dまでの間は、全て、一般的なファイバチャネルプロトコルに準拠しているため、FC-AIに対応する機器を利用して接続することが可能である。

【0078】以上説明したように、本実施の形態によれば、リモートディスクサブシステム7に対してリモートコピーを行うマスターディスクサブシステム1に複数のホストコンピュータが接続される場合、各ホスト単位に設定した個別比率閾値32や合計比率閾値40等の比率情報を用いて、共有資源であるキャッシュメモリ11内のサイドファイルを各ホストの重要度等に応じて効率的に配分して使用できるため、ユーザが意図した優先度の通りに、各ホストコンピュータのデータをマスターディスクサブシステム1からリモートディスクサブシステム7側へと確実にリモートコピーできる。

【0079】従って、たとえば、マスターディスクサブシステム1において、特定の1つのホストコンピュータからの大量なデータ書き込みが発生しても、他の重要度の高いホストコンピュータの入出力処理は影響されことなく、当該重要度の高いホストコンピュータからのアクセスを受け付けながら、リモートコピーの動作を継続できる。

【0080】また、本実施の形態の場合には、メインフレーム系チャネルプロトコルのCKD形式等のデータ形式を、オープン系のファイバチャネルプロトコル・チャネルに準拠したデータ形式に変換した後に、ディスクアレイ12への格納およびリモートディスクサブシステム7へのリモートコピーを実行するので、マスターディスクサブシステム1とリモートディスクサブシステム7の接続インタフェースとして、一種類のリモートコピー用ファイバチャネル6のみを用いるだけで、ファイバチャネルプロトコル・チャネルに接続するオープン系システムだけでなく、メインフレームホストコンピュータで扱

うCKDフォーマットデータについても同様の制御にてリモートコピーが可能であり、低コストで効率的なりモートコピーシステムの構築が可能となる。

【0081】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0082】たとえば、上述の実施の形態の説明では、一例として、自動計測したサイドファイル使用済み容量（個別使用量33、合計使用量34）と、SVP14を経由してユーザーが予め設定した合計比率閾値40、個別比率閾値32とを比較して、データ流入制限の指標としているが、ユーザー設定に限らず、たとえば長期間に渡って自動計測したサイドファイル使用量等の統計的な情報から上述の各種閾値の指標を定める制御論理を記憶制御装置等に付加することも本発明に含まれる。

【0083】

【発明の効果】本発明によれば、記憶サブシステムに接続される複数の上位装置の各々の重要度に応じた優先度にて、複数の記憶サブシステム間でのライトデータのリモートコピーを行うことができる、という効果が得られる。

【0084】また、本発明によれば、記憶サブシステムに接続される複数の上位装置のうち、特定の上位装置によるデータ書込負荷の増大に影響されることなく、複数の記憶サブシステム間での効率的なりモートコピーを実現することができる、という効果が得られる。

【0085】また、本発明によれば、複数の上位装置の各々と記憶サブシステム間の多様な接続インタフェースを、データのリモートコピーを行う複数の記憶サブシステム間の接続インタフェースに統一することで、効率的なりモートコピーシステムの構築ができる、という効果が得られる。

【図面の簡単な説明】

【図1】本発明の一実施の形態である記憶サブシステムの制御方法を実施する記憶サブシステムを用いた情報処理システムの構成の一例を示す概念図である。

【図2】本発明の一実施の形態である記憶サブシステムの内部構成の一例をより詳細に例示した概念図である。

【図3】本発明の一実施の形態である記憶サブシステムにて用いられる制御情報の一例を示す概念図である。

【図4】本発明の一実施の形態である記憶サブシステムの作用の一例を示す概念図である。

【図5】本発明の一実施の形態である記憶サブシステムにおけるリモートコピーの作用の一例を示す概念図である。

【図6】本発明の一実施の形態である記憶サブシステムにおけるデータフォーマットの変換方法の一例を示す概念図である。

【図7】本発明の一実施の形態である記憶サブシステムにおけるファイバチャネルプロトコルによる記憶サブシステム間の接続方法の変形例を示す概念図である。

【図8】本発明の一実施の形態である記憶サブシステムの作用の一例を示すフローチャートである。

【図9】本発明の一実施の形態である記憶サブシステムの作用の一例を示すフローチャートである。

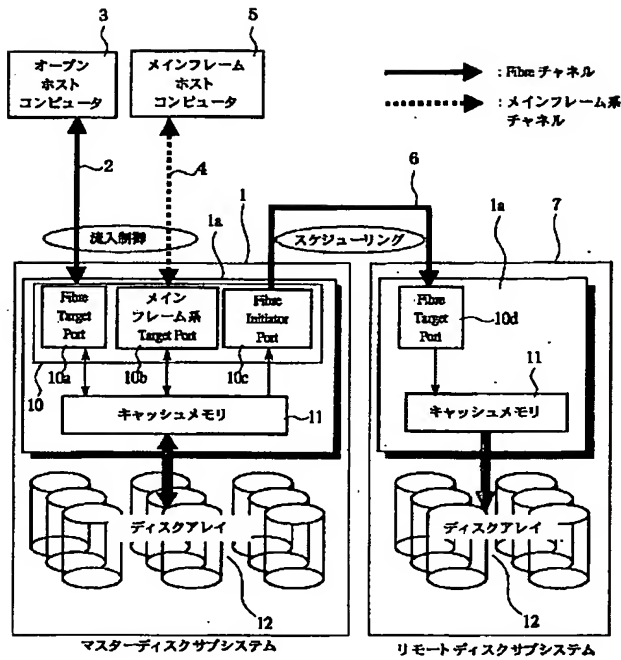
【図10】本発明の一実施の形態である記憶サブシステムの作用の一例を示すフローチャートである。

【符号の説明】

1…マスターディスクサブシステム（第1の記憶サブシステム）、1a…記憶制御装置、2…ファイバチャネル（第1のインタフェース）、3…オープンホストコンピュータ、4…メインフレーム系チャネル（第1のインタフェース）、5…メインフレームホストコンピュータ、6…リモートコピー用ファイバチャネル（第2のインタフェース）、7…リモートディスクサブシステム（第2の記憶サブシステム）、10…入出力ポート、10a…ファイバチャネルターゲットポート、10b…メインフレーム系チャネルターゲットポート、10c…ファイバチャネルイニシエータポート、10d…ファイバチャネルターゲットポート、11…キャッシュメモリ、11-1…共用メモリ、12…ディスクアレイ、12a…ディスクアダプタ、13…バス、14…サービスプロセッサ（SVP）、15…外部接続インタフェース、30…サイドファイル管理テーブル、31…ポート番号、32…個別比率閾値（第1の閾値）、33…個別使用量、34…合計使用量、40…合計比率閾値（第2の閾値）、41…サイドファイル容量、50…キャッシュ管理テーブル、51…セグメント番号、52…マスタダミーフラグ、53…リモートダミーフラグ、54…LRU管理情報、60…ファイバチャネルハブ、70…ファブリック、71…ファイバチャネルスイッチ。

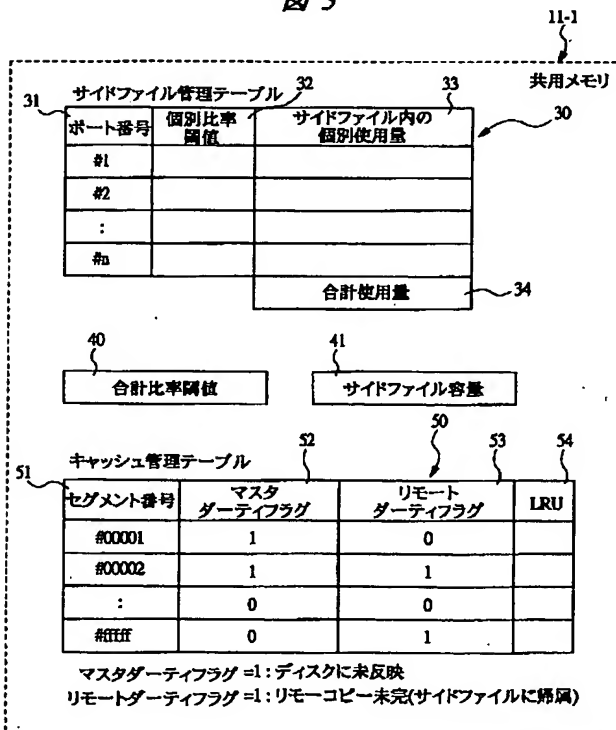
【図1】

図 1



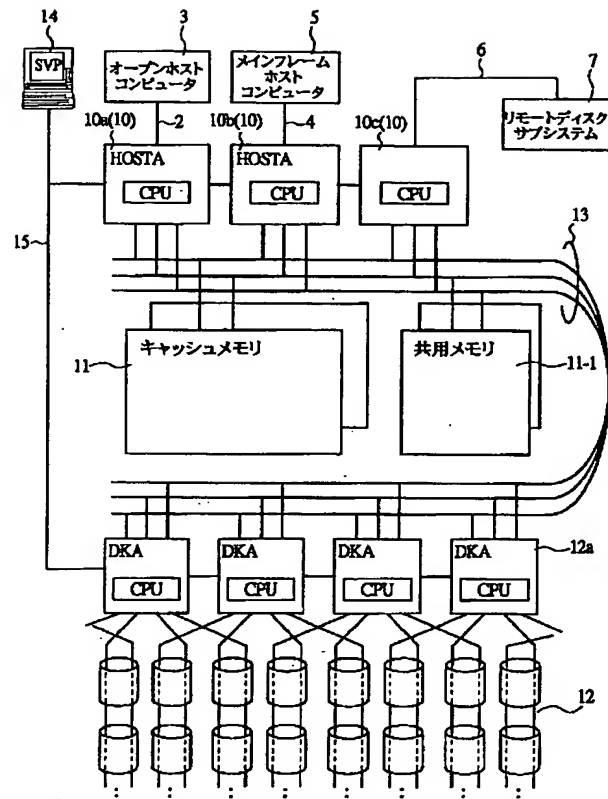
【図3】

図 3



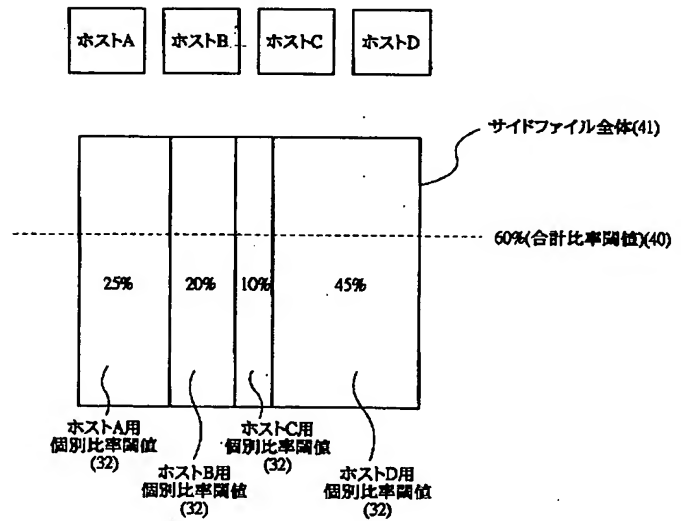
【図2】

図 2

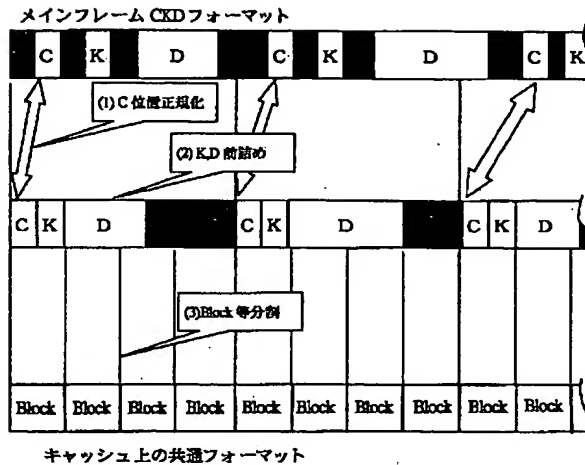


【図4】

図 4

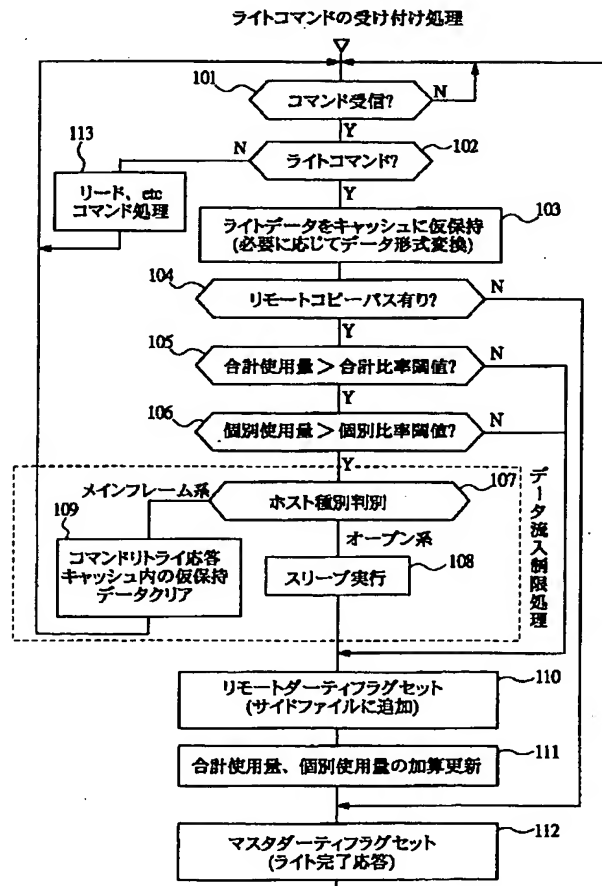


【図 6】



【图 8】

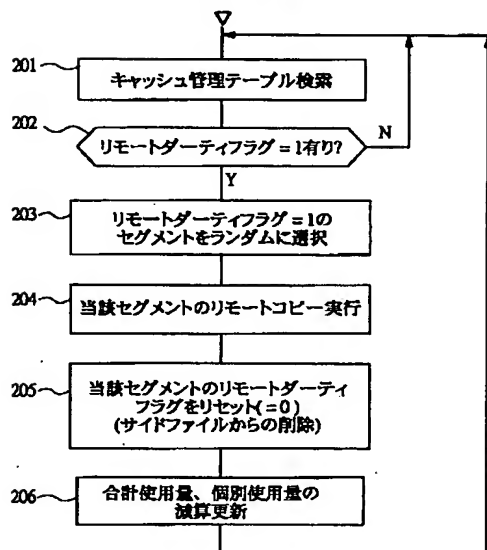
**8**



【図9】

図 9

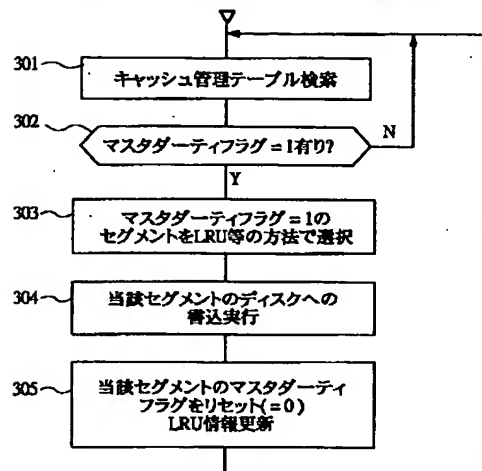
リモートコピー処理



【図10】

図 10

キャッシュメモリからディスクへのデータ書込処理



フロントページの続き

(51) Int. Cl.<sup>7</sup>

G 0 6 F 12/08

識別記号

5 4 1

5 4 3

5 5 7

12/16

3 1 0

13/10

3 4 0

F I

G 0 6 F 12/08

タームコード (参考)

5 4 1 C

5 4 3 C

5 5 7

12/16

3 1 0 J

13/10

3 4 0 B

F ターム (参考) 5B005 KK15 LL11 MM11 MM22 NN12

5B014 EB05 GC27

5B018 GA04 HA04 KA03 MA12

5B065 BA01 CA11 CC08 CE12 CE21

EA35 ZA15

5B082 DE07 FA02 HA05